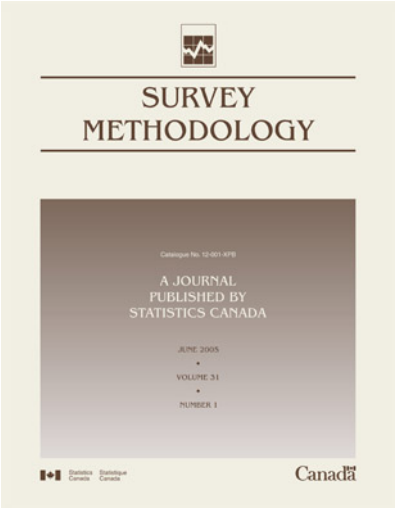
Catalogue no. 12-001-XIE

Survey Methodology

June 2004



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Business Survey Methods Division

Survey Methodology

June 2004

Published by authority of the Minister responsible for Statistics Canada

April 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 30, No. 1, pp. 45-55
Statistics Canada, Catalogue No. 12-001

Properties of the Weighting Cell Estimator Under a Nonparametric Response Mechanism

D. Nóbrega Da Silva and Jean D. Opsomer [1]

Abstract

The weighting cell estimator corrects for unit nonresponse by dividing the sample into homogeneous groups (cells) and applying a ratio correction to the respondents within each cell. Previous studies of the statistical properties of weighting cell estimators have assumed that these cells correspond to known population cells with homogeneous characteristics. In this article, we study the properties of the weighting cell estimator under a response probability model that does not require correct specification of homogeneous population cells. Instead, we assume that the response probabilities are a smooth but otherwise unspecified function of a known auxiliary variable. Under this more general model, we study the robustness of the weighting cell estimator against model misspecification. We show that, even when the population cells are unknown, the estimator is consistent with respect to the sampling design and the response model. We describe the effect of the number of weighting cells on the asymptotic properties of the estimator. Simulation experiments explore the finite sample properties of the estimator. We conclude with some guidance on how to select the size and number of cells for practical implementation of weighting cell estimation when those cells cannot be specified *a priori*.

Key Words: Finite population asymptotics; Quasi-randomization inference; Weighting cell selection.

1. Introduction

Item and unit nonresponse occur in almost all large-scale surveys, and proper estimation techniques need to account for it. While item nonresponse is often dealt with through imputation, unit nonresponse is most often accounted for through weighting adjustments. Cell weighting adjustments for nonresponse have been applied since at least the 1950s in survey estimation, *e.g.*, U.S. Bureau of the Census (1963, page 53), and continue to be widely used in practice today, because they have intuitive appeal and are relatively easy to implement in practice. Reviews of common weighting procedures are given in Kalton (1983) and Kalton and Kasprzyk (1986). A number of authors have studied the properties of the weighting cell estimator under a variety of theoretical frameworks. Oh and Scheuren (1983) derive the mean and variance of the weighting cell estimator under simple random sampling, conditional on the sample size and the number of respondents in each cell. See also Kalton and Maligalig (1991). Särndal, Swensson and Wretman (1992, page 578) use the term "response homogeneity group" for cells in which the nonresponse is assumed to be constant, and derive the properties of the resulting weighting cell estimator for general designs. The recently introduced *fully efficient fractional imputation* (FEFI) of Kim and Fuller (1999) can also be expressed as a weighting cell estimator, and these authors derive its model properties under the assumption that the variables are independent and identically distributed (iid) within each cell.

While the specific assumptions vary, a common thread among all these results is that the weighting cells are correctly specified, in the sense that units within each cell are indeed fully "exchangeable" (the precise definition of this term depends on the framework selected: equal response probabilities for randomization-based inference, or iid observations for model-based inference). In the terminology of Little and Rubin (2002, Chapter 1), this is the case of observations *missing at random* (MAR), where auxiliary information (*i.e.*, cell membership in this case) can be used to correct the inference for the nonresponse.

In this article, we depart from this framework. We will assume that the response mechanism depends on a known continuous auxiliary variable, but the exact functional form of this relationship is left almost completely unspecified (details on this *nonparametric response mechanism* are provided in the next section). Knowledge of such a variable could be used to construct more sophisticated nonresponse adjustments such as *propensity weighting* (Cassel, Särndal and Wretman (1983), Little (1986), and Da Silva and Opsomer (2003)) or post-stratification, but we will instead limit our use of this auxiliary variable to the division of the population into weighting cells. Our primary goal with this approach is to study the robustness of the popular weighting cell estimator to model misspecification, and in particular, the effect of the number of cells. Hence, in contrast to the approach of the authors discussed above, the weighting cells are used as a practical way to construct an survey estimator, but they will not be assumed as part of the

1. D. Nacimento Da Silva, Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal, RN 59072-970, Brazil. E-mail: damiao@ccet.ungrn.br; Jean D. Opsomer, Department of Statistics, Iowa State University, Ames IA 50011, U.S.A. E-mail: jopsomer@iastate.edu.

statistical framework. This is similar to the "adjustment by subclassification" idea proposed by Cochran (1968) for removing the bias due to a continuous covariate in observational studies.

We will study the properties of the estimator under *quasi-randomization*, a term used by Oh and Scheuren (1983) to denote joint inference under the sampling design and the response mechanism. The asymptotic properties of the estimator will be established by embedding the finite population and the corresponding sampling design and response mechanism in a sequence of such populations and random mechanisms, as will be explained in later sections. This asymptotic framework is very similar to that advocated by Hansen, Madow and Tepping (1983) and used in Isaki and Fuller (1982), among others.

The remainder of this paper is as follows. In section 2, we introduce the notation and framework for the sampling design and the nonresponse model, and discuss the weighting cell estimator. In the following section, we derive the asymptotic design properties of the estimator. In section 4, we report on a simulation study to examine the practical behavior of the estimator, compare its practical behavior with that predicted by the asymptotic theory, and provide some guidance on the choice of the weighting cells.

2. The Weighting Cell Estimator

Before describing the weighting cell estimator, we introduce our survey design framework and the response generating mechanism. We consider a population $U = \{1, 2, ..., N\}$, where N is finite and known. For every element i in U, let $\boldsymbol{Y}_i = (Y_{1,i}, Y_{2,i}, ..., Y_{p,i})$ be the associated vector of values of p characteristics of interest, $Y_1, Y_2, ..., Y_p$. Likewise, let $\boldsymbol{X}_i = (X_{1,i}, X_{2,i}, ..., X_{q,i})$ be the vector of values of q auxiliary variables, $X_1, X_2, ..., X_q$, corresponding to the i^{th} unit, $i \in U$. We assume that $\boldsymbol{X}_i$ is known $\forall i \in U$. If $p = 1$, we denote $\boldsymbol{Y}_i$ by Y_i and, for $q = 1$, X_i is used to denote $\boldsymbol{X}_i$. Let s represent a sample drawn from U according to some sampling design $p(\cdot)$. This sampling design $p(\cdot)$ is chosen by the survey sampler and may be based on information available in the $\boldsymbol{X}_i$, $i \in U$.

The goal of the sample survey is to estimate unknown population quantities such as the population mean or total, or a function of these quantities. To simplify the presentation, we will focus on the estimation of the population total of the $\boldsymbol{Y}_i$,

$$\boldsymbol{t}_y = \sum_U \boldsymbol{Y}_i.$$

When there is no nonresponse, this quantity will be estimated by a sample-based estimator of the form

$$\hat{\boldsymbol{t}}_y = \sum_s w_i \boldsymbol{Y}_i = \sum_U w_i \boldsymbol{Y}_i I_i \qquad (1)$$

where the w_i, $i \in s$, are the sampling weights and I_i is an indicator for whether the i^{th} unit is in the sample or not. In this article, we will assume that the sampling weights are the inverse of the inclusion probabilities, or $w_i = \pi_i^{-1}$, with $\pi_i = \Pr(i \in s)$, so that the estimator (1) is the classical Horvitz-Thompson estimator (Horvitz and Thompson 1952). Also, let $\boldsymbol{I} = (I_1, I_2, ..., I_N)^T$ represent the vector of inclusion indicators for the population.

In the context of nonresponse, it is convenient to assume that each unit in the population is either a *respondent* or a *nonrespondent* for the variable of interest $\boldsymbol{Y}$. Consider the vector $\boldsymbol{R} = (R_1, R_2, ..., R_N)^T$, where R_i indicates if the i^{th} unit is a respondent or not. The distribution of $\boldsymbol{R}$ is called the *response mechanism*. In analogy to the definition of the sample s, we use $r \subseteq U$ to denote the (realized) set of respondents in the population, *i.e.*, those elements for which $R_i = 1$. Since the distribution of r and $\boldsymbol{R}$ is typically unknown and can in principle depend on the realized value of $\boldsymbol{I}$ as well as on the $\boldsymbol{Y}$, we need to assume a model for the response mechanism. When this assumed model is used to develop an estimator for a population quantity, the properties of this estimator become dependent on the response model. Hence, a misspecified model for $\boldsymbol{R}$ has the potential to cause significant and difficult to measure bias in both the estimator and its associated measures of precision. To avoid this problem, we will keep the response mechanism quite general in this article. Specifically, we will assume that the R_i are independent Bernoulli variables with

$$\Pr\{R_i = 1 \mid \boldsymbol{I}, \boldsymbol{Y}\} = \varphi_i, \; 0 < \varphi_i \leq 1, \; \forall i \in U,$$

and that the φ_i can be written as $\varphi_i = \varphi(\boldsymbol{X}_i)$, with $\varphi(\cdot)$ a continuous and differentiable but otherwise unspecified function of the $\boldsymbol{X}_i$. Note that this includes the uniform response mechanism, where $\varphi_i \equiv \varphi$ for all $i \in U$, as a special case.

When some of the selected elements do not respond, the estimator (1) can no longer be computed, and an estimator that includes a nonresponse adjustment is required. In this article, we are using the weighting cell estimator for this purpose. For simplicity, we will describe the situation in which both the $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are univariate variables, but the approach can be generalized to the multi-dimensional case. Let $s_r = s \cap r$ represent the subset of the selected elements that actually respond to the survey.

Let U_g, $g = 1, ..., G$, represent G groups obtained by dividing the population into groups based on the values of the known auxiliary variable X. Specific implementations might generate groups of equal size, or divide the range of

X into equal-length intervals. We shall leave the implementation unspecified for now, and state some general assumptions about G and the size of the groups in the next section. Note that we are considering the groups as fixed with respect to the sampling design and the response mechanism, which excludes the situation in which groups are formed based on the *observed* sample values $\{X_i : i \in s\}$. This was done primarily to simplify the theoretical derivations, and is similar to the approach of Särndal *et al.* (1992) and Kim and Fuller (1999), among others.

Let $s_g = s \cap U_g$ be the portion of the sample that falls in group g, and define similarly $s_{r,g} = s_r \cap U_g$. The weighting cell estimator is defined as

$$\hat{t}_{\mathrm{WC}} = \sum_{g=1}^{G} \left(\frac{\sum_{s_g} w_i}{\sum_{s_{r,g}} w_i} \right) \sum_{i \in s_{r,g}} w_i Y_i. \qquad (2)$$

From this expression, is it easy to see that in each group, the estimator of the group total is ratio-adjusted by the inverse of the weighted proportion of respondents in the cell. This estimator is also the FEFI estimator of Kim and Fuller (1999). The properties of this estimator will be studied in next section.

3. Properties Under Quasi-Randomization

3.1 Asymptotic Framework and Assumptions

The quasi-randomization properties of the weighting cell estimator will be studied in the usual finite population asymptotic context, in which the population U is treated as an element in an increasing sequence $U_1, U_2, ..., U_v$ with $v \to \infty$, with a corresponding sequence of sampling designs $p_v(\cdot)$ (see Isaki and Fuller (1982) for an early example of this framework). Let N_v be the size of the v^{th} population with $N_v > N_{v-1}$, let $\boldsymbol{Y}_v = (Y_1, Y_2, ..., Y_{N_v})^T$ denote the set of values of the characteristic of interest, Y, associated with U_v, and similarly, $\boldsymbol{X}_v = (X_1, X_2, ..., X_{N_v})^T$. We assume that $\boldsymbol{X}_v$ is known. For each v, a sample of size n_v $(n_v \geq n_{v-1})$ is selected from U_v, according to a sampling design $p_v(\cdot)$. As before, let $\boldsymbol{I}_v = (I_1, I_2, I_{N_v})^T$ be the corresponding sample inclusion vector. We will denote the K^{th} order central moment of the sample membership indicators $I_{i_1}, ..., I_{i_k}$ by

$$\Delta_{i_1,...,i_K} = E\left(\prod_{k=1}^{K} (I_{i_k} - \pi_{i_k}) \right). \qquad (3)$$

It is assumed that U_v can be divided into G_v $(G_v \geq G_{v-1})$ mutually exclusive and exhaustive groups, U_g, $g = 1, ..., G_v$. These groups are constructed by sorting the population according to their X values and dividing the population into G_v groups. We will assume that there are at least G_v distinct values among the elements of $\boldsymbol{X}_v$. Let N_g represent the number of elements in U_g.

As mentioned in the previous section, we are treating the groups as fixed with respect to the population. The problem created by this approach is that in general, there is a non-zero chance of obtaining a group without any respondents. We solve this problem by adding a small constant in the denominators in each of the groups, or

$$\hat{t}^{*}_{\mathrm{WC}} = \sum_{g=1}^{G_v} \left(\frac{\sum_{s_g} w_i}{\max\left(\sum_{s_{r,g}} w_i, N_g\, G_v\, n_v^{-1} \right)} \right) \sum_{i \in s_{r,g}} w_i\, Y_i. \qquad (4)$$

Hence, the difference between $\hat{t}^{*}_{\mathrm{WC}}$ and $\hat{t}_{\mathrm{WC}}$ in (2) is asymptotically negligible. This is similar to what is often done in practice to avoid overly large weights in ratio estimation.

Fuller and Kim (2003) give the limiting distribution of the FEFI estimator under the assumption that the response probabilities are constant within these cells. We will study the case where the response probabilities are a smooth function of an auxiliary variable and the number of cells are allowed to vary. Let $\mathbf{R}_v = (R_1, R_2, ..., R_{N_v})^T$ be the response indicator vector for the v^{th} population. We assume that the distribution of $\mathbf{R}_v$ satisfies the *nonparametric response mechanism* assumptions, specified as follows:

(R1) $R_1, R_2, ..., R_{N_v}$ are independent random variables,
(R2) $\Pr\{R_i = 1 | \mathbf{I}_v, \boldsymbol{Y}_v\} = \varphi_i$, $\forall i \in U_v$,
(R3) $\varphi_i = \varphi(X_i)$ $\forall i \in U_v$, where $\varphi(\cdot)$ is differentiable with bounded first derivative, and the $X_i \in [x_m, x_M]$, with x_m, x_M fixed constants and $x_m < x_M$.

The remaining assumptions are technical conditions that will be used extensively in the proofs. We assume that there are positive constants $\lambda_1, \lambda_2, ..., \lambda_9$ such that:

(A1) $\lambda_1 < N_v\, n_v^{-1}\, \pi_i < \lambda_2 < \infty$, $\forall i \in U_v$, and $n_v\, N_v^{-1} \to \pi \in (0,1)$, as $v \to \infty$;

(A2) For distinct $i_1, ..., i_K \in U_v$, $K = 2, 3, ..., 8$,

$$\left|\Delta_{i_1,...,i_K}\right| \leq \begin{cases} \left(\prod_{k=1}^{K}(N-k+1)\right)^{-1} n_v^{K/2} \lambda_3, & \text{if K is even} \\ \left(\prod_{k=1}^{K}(N-k+1)\right)^{-1} n_v^{(K-1)/2} \lambda_4, & \text{if K is odd} \end{cases}$$

(A3) $\lim_{v\to\infty} \frac{1}{N_g} \sum_{i \in U_g} \varphi_i = \varphi^{*}_g, \forall g = 1, 2, ..., G_v$ and $v \geq 1$;

(A4) $\max_{i \in U_v} |Y_i| \leq \lambda_5$;

(A5) $\lambda_6 < \min_{i \in U_v} \varphi_i \leq 1$;

(A6) $\lambda_7 G_v^{-1} \le N_g N_v^{-1} \le \lambda_8 G_v^{-1}, \forall g = 1, 2, ..., G_v$;

(A7) $1 \le G_v \le n_v^{\gamma} \lambda_9$, with $0 \le \gamma \le 1/2$.

Assumptions (A1)–(A2) imply that, asymptotically, the sampling design is "well behaved," in the sense that the moments of the sample membership indicators are of the same order of magnitude as those in simple random sampling without replacement. This is a common assumption in finite population asymptotic theory. (A1) also requires that the sampling fraction converges to a constant in the interval (0, 1). The boundedness assumption (A4) on the observations will significantly simplify the proofs for some of the theorems in the article, and could be relaxed to the existence of bounded moments if desired. Similarly, some technical regularity conditions are required to avoid degenerate response mechanisms: (A3) provides that the limit for the average response probability in a cell exists, and (A5) excludes the situation in which some units might have $\varphi_i = 0$. Finally, assumptions (A6) and (A7) on the weighting cells require that all the cells grow at a similar rate, and that the total number of cells does not increase "too fast" relative to the sample size.

3.2 Main Results

The approach we will use in the study of the properties of the weighting cell estimator follows that commonly used in the study of finite population estimators. First, we show the asymptotic equivalence between the non-linear weighting cell estimator and a "linearized" approximation. Next, we derive the mean squared error properties of the linearized estimator and consider those as the asymptotic properties of the weighting cell estimator or, more precisely, the properties of the asymptotic distribution of the weighting cell estimator. See, for instance, Särndal *et al.* (1992, Chapter 5) for a description of this approach.

The following theorem formally states our first results. The proof is in the appendix.

Theorem 3.1. *Consider the sequence of populations* $\{U_v: v \ge 1\}$. *Assume that for each* v, *a probabilistic sample of fixed size* n_v ($n_v \ge n_{v-1}$) *is selected from* U_v *according to sampling design* $p_v(\cdot)$, *and that the response mechanism satisfies the conditions* $(R1)-(R2)$. *Finally, assume that* $(A1)-(A7)$ *hold. Then, the estimator* $\hat{t}_{\mathrm{WC}}^*$ *is asymptotically equivalent to a linearized random variable* $\tilde{t}_{\mathrm{WC}}$, *in the sense that*

$$\frac{1}{N_v}(\hat{t}_{\mathrm{WC}}^* - \tilde{t}_{\mathrm{WC}}) = O_p(G_v n_v^{-1}). \quad (5)$$

The bias and variance of $\tilde{t}_{\mathrm{WC}} / N_v$ *are given by*

$$E\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right) - \bar{Y}_v = \frac{1}{N_v}\sum_{g=1}^{G_v}\sum_{U_g}\left(\frac{\varphi_i - \bar{\varphi}_g}{\bar{\varphi}_g}\right)(Y_i - \tilde{Y}_g) \quad (6)$$

and

$$\mathrm{Var}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right) = \frac{1}{N_v^2}\sum_{g=1}^{G_v}\sum_{g'=1}^{G_v}\left[\sum_{U_g}\sum_{U_{g'}}\Delta_{ij}\tilde{Y}_{ig}\tilde{Y}_{jg'}\right] + \frac{1}{N_b^2}\sum_{g=1}^{G_v}\sum_{U_g}\pi_i^{-2}\frac{\varphi_i(1-\varphi_i)}{\bar{\varphi}_g^2}(Y_i - \tilde{Y}_g)^2, \quad (7)$$

where

$$\bar{\varphi}_g = \frac{1}{N_g}\sum_{U_g}\varphi_i,\ \bar{Y}_g = \frac{1}{N_g}\sum_{U_g}Y_i,\ \tilde{Y}_g = \frac{\sum_{U_g}\varphi_i Y_i}{\sum_{U_g}\varphi_i}$$

and

$$\tilde{Y}_{ig} = \frac{\varphi_i(Y_i - \tilde{Y}_g) + \bar{\varphi}_g\tilde{Y}_g}{\pi_i\bar{\varphi}_g},$$

$$\forall i \in U_g \text{ and } \forall g = 1, 2, ..., G_v.$$

Remark 1. The asymptotic equivalence between $\hat{t}_{\mathrm{WC}}^*$ and $\tilde{t}_{\mathrm{WC}}$ depends on the number of groups G_v, with a faster convergence rate achieved when G_v grows more slowly. The intuition behind this result is that the goodness of the linear approximation depends on how well the true cell ratio response adjustments φ_g^* are estimated by the sample-based estimators $\sum_{s_{r,g}} w_i / \sum_{s_g} w_i$. Since the cell ratios will be better estimators as the sample size grows larger, this would argue that G_v should be chosen to be small, which corresponds to the current practice in applications of weighting cell estimation. However, as will be shown below, the MSE properties of $\tilde{t}_{\mathrm{WC}}$ under the nonparametric response mechanism improve as G_v gets larger. A more detailed discussion of the selection of the number of groups will be provided after Theorem 3.2 below and in section 4.

Remark 2. The results in Theorem 3.1 depend on the population groups U_g, $g = 1, ..., G_v$ and on the φ_i, $i \in U_v$, but do not rely on the fact that the response probabilities are a smooth function of the auxiliary variable X. Hence, the explicit expressions for the asymptotic bias and variance can be used to derive results for other response mechanisms that follow (R1)–(R2). In particular, results for the *response homogeneity group* model (see Särndal *et al.* 1992, page 577) follow directly from Theorem 3.1. This is also the model studied by Fuller and Kim (2003). Under that model, one assumes that $\varphi_i \equiv \varphi_g$ for all

$i \in U_g$, $g = 1, ..., G$, and it can easily be shown that the bias of $\tilde{t}_{\mathrm{WC}}$ is 0 and its variance is

$$\mathrm{Var}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right) = \mathrm{Var}\left(\frac{\hat{t}_y}{N_v}\right) + \frac{1}{N_v^2}\sum_{g=1}^{G_v}\frac{1-\varphi_g}{\varphi_g}\sum_{U_g}\pi_i^{-2}(Y_i - \bar{Y}_g)^2.$$

The first term in the variance is the variance of the estimator without nonresponse, and the second term represents the variance inflation caused by the nonresponse under a homogeneous within-cell response mechanism.

The following corollary follows directly from Theorem 3.1 and Fuller (1996, Theorem 5.2.1). A proof is given in the appendix.

Corollary 3.1. *Under the conditions of Theorem 3.1 with* $\gamma < 1/2$ *in* (*A7*), *for any sampling design* $p_v(\cdot)$ *such that*

$$n_v^{1/2}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v} - \bar{Y}_v - B_v\right) \xrightarrow{L} N(0, V),$$

where B_v *corresponding to the bias of* $\tilde{t}_{\mathrm{WC}}/N_v$ *given in Theorem 3.1 and*

$$V \equiv \lim_{v\to\infty} n_v \mathrm{Var}(\tilde{t}_{\mathrm{WC}}/N_v) \in (0, \infty),$$

then

$$\left[\mathrm{Var}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right)\right]^{-1/2}\left(\frac{\tilde{t}^*_{\mathrm{WC}}}{N_v} - \bar{Y}_v - B_v\right) \xrightarrow{L} N(0, 1).$$

Corollary 3.1 states that, whenever the linearized estimator $\tilde{t}_{\mathrm{WC}}$ achieves asymptotic normality, then so does $\hat{t}^*_{\mathrm{WC}}$. Since $\tilde{t}_{\mathrm{WC}}$ can be written as a classical expansion estimator of the form (1), this result is quite general.

Under the nonparametric response mechanism described in (R1) – (R3), it is possible to describe the affect of the number of groups G_v on the asymptotic bias and variance of $\hat{t}^*_{\mathrm{WC}}$. The next theorem gives the asymptotic rates for the bias and variance, and is proven in the appendix.

Theorem 3.2. *Assume that* (*R*3) *and the conditions of Theorem* 3.1. *Then*,

$$E\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right) - \bar{Y}_v = O\left(\frac{1}{G_v}\right)$$

and

$$\mathrm{Var}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right) = O\left(\frac{1}{n_v}\right) + O\left(\frac{1}{n_v G_v}\right).$$

Remark 3. Theorem 3.2 shows that both the asymptotic bias and variance of the weighting cell estimator $\hat{t}^*_{\mathrm{WC}}$ become smaller as the number of groups G_v increases. An intuitive explanation of that fact is that the approximation of the function $\varphi_i = \varphi(X_i)$ by the step function $\varphi_i = \varphi^*_g$ improves as the number of cells increases. The asymptotic variance has a term that is independent of G_v. This "residual variance" is due to the inherent variability of the sampling design and the response mechanism, and cannot be reduced by changing G_v.

Remark 4. As noted in Remark 1, constructing a good linear approximation $\tilde{t}_{\mathrm{WC}}$ requires G_v to be small, while Theorem 3.2 states that the MSE of $\tilde{t}_{\mathrm{WC}}$ is minimized by taking G_v as large as possible. Taken together, this can be interpreted to mean that, once the sample size in every cell is sufficiently large to obtain a "valid" ratio estimator for the average cell response probability φ^*_g, it is preferable to increase the number of cells than to increase the sample size per cell. The simulation experiments discussed in section 4 will further explore this recommendation.

The following corollary follows directly from Corollary 3.1, Theorem 3.2, and Chebyshev's inequality, and establishes the consistency of the weighting cell estimator under the nonparametric response mechanism.

Corollary 3.2. *Under the conditions of Theorem* 3.2, $\hat{t}^*_{\mathrm{WC}}$ *is a consistent estimator for* t_y, *in the sense that for any* $\in > 0$,

$$\Pr\left(\left|\frac{\tilde{t}^*_{\mathrm{WC}} - t_y}{N_v}\right| > \in\right) \to 0, \; v \to \infty.$$

Remark 5. As Corollary 3.2 shows, as long as a variable X can be found that is sufficiently related to the nonresponse, in the sense of assumptions (R1) – (R3), construction of weighting cells does not require knowledge of homogeneous response probability cells in order to construct a consistent estimator. However, as discussed in Remarks 1 and 4, the choice of the number of cells still has an effect on the properties of the estimator.

Remark 6. Assumption (R3) can easily be relaxed to allow for a small number of points of discontinuity in both $\varphi(\cdot)$ and its first derivative. A "small" number can mean that the number is either fixed as $v \to \infty$ or increases at a rate slower than G_v. This would make it possible to account for situations such as stratified designs or the presence of domains within U_v. The present theory can be extended

directly to these situations, if the values for the variable X fall in non-overlapping segments for the different strata or domains.

4. Simulation Experiments

4.1 Description of the Experiment

In order to investigate the practical implications of the results of section 3, we carried out a Monte Carlo experiment on a fixed population of $N = 3{,}000$ units. We consider the case of one covariate, X whose population values are generated as:

$$X_1,\ X_2,\ ...,\ X_N \ \sim \text{ i.i.d. } U(0, 1),$$

and two different variables of interest, Y_1 and Y_2. We are interested in evaluating the effects of (1) the (model) relationship between Y and X, (2) the response mechanism $\varphi(X)$, (3) the sample size n and (4) the number of cells G, on the bias and on the mean square error of the $\hat{t}_{\text{WC}}$ estimator. Since our theoretical results rely on the approximation of $\hat{t}_{\text{WC}}$ (or $\hat{t}^{*}_{\text{WC}}$) by a linearized estimator $\tilde{t}_{\text{WC}}$, we will also compare the behavior of $\hat{t}_{\text{WC}} / N_v$ and $\tilde{t}_{\text{WC}} / N_v$ as estimators of the population mean, $\bar{Y}_v = N_v^{-1} \sum_U Y_i$. Finally, we compare $\hat{t}_{\text{WC}} / N_v$ to the "naive" estimator of the mean, which is defined for the variable Y as:

$$\bar{y}_r = \frac{\sum_{i \in s_r} w_i\ Y_i}{\sum_{i \in s_r} w_i},$$

corresponding to a ratio adjustment of the respondent sample to the original sample. This estimator is appropriate under the assumption of uniform response mechanism or, to use the terminology of Little and Rubin (2002, chapter 1), when observations are *missing completely at random* (MCAR). Note that $\bar{y}_r$ is equivalent to the weighting cell estimator with a single cell.

The levels of the four factors used in the experiment are given in Table 1. The "levels" of the variable Y correspond to two populations of independent values. The variable Y_1 was generated as $N(40,\ 58)$, truncated to -3 to $+3$ standard deviations, corresponding to the "white noise" case. The variable Y_2 is related to X and was generated through the linear model $Y_2 = 27.12 + 26.06 X + \varepsilon$, where $\varepsilon \sim N/(0, 9)$. The population mean and variance for the two variables were, respectively, (39.9, 55.3) for Y_1, and (40.0, 63.9) for Y_2.

The four levels of the response mechanisms contain two different scenarios regarding the response probabilities: constant (C1, C2), and linearly related to X (L1, L2). The response probabilities are:

$$-\varphi_{C1}(X) = 0.5$$

$$-\varphi_{C2}(X) = 0.8$$

$$-\varphi_{L1}(X) = 0.20 + 0.60X$$

$$-\varphi_{L2}(X) = 0.65 + 0.30X$$

The levels of the linear response mechanisms were chosen so that the average probabilities (over X) were approximately equal to 0.5 and 0.8, respectively.

Table 1
Overview of Factors in the Simulation Experiment

Factor	Levels
Y variable	$Y_1,\ Y_2$
Response mechanism $\varphi(\cdot)$	C1, C2, L1, L2
Sample size n	200, 500
Number of cells G	2, 3, 5, 8

For a given G, the groups were created by dividing the range of X into G equal segments and assigning the element i to the group g if the value X_i was in the g^{th} segment, $i = 1, 2, ..., N$ and $g = 1, 2, ..., G$. The simulations were carried out through a completely randomized factorial experiment $2 \times 4 \times 2 \times 4$. For each combination of the levels of the factors in Table 1, $B = 5{,}000$ independent realizations of the vector indicator of responses, $\boldsymbol{R} = (R_1, R_2, ..., R_N)^T$, were generated according to the corresponding response mechanism. For each one of such realizations, a simple random sample (without replacement and of size n), s, was selected from the overall population. Within each selected sample, the respondents were the values of $i \in s$ such that $R_i = 1$.

This procedure could in principle lead to a group not containing any sampled and responding element, in which case weighting cell estimator (ignoring the adjustment in (4)) cannot be computed. If that happened, the realization was discarded and a new sample drawn from the population. Out of the 5,000 repetitions for each combination of factors, this happened 13 times in the factor combination $(Y_1,\ \varphi_{L1},\ 200, 8)$ and 15 times with $(Y_2,\ \varphi_{L1},\ 200, 8)$. It did not occur with any of the other factor combinations. Hence, the number of samples discarded was very small and this has a negligible effect on the simulation results.

With $n = 200$ and $G = 8$, we expect approximately 25 sampled elements in each cell, to be further reduced by the nonresponse. Since the estimator relies on ratio estimation in each cell, we judged this to be a reasonable lower bound on the number of observations per cell to consider in the simulations. In practice, a number of procedures could be used when groups have too few elements, such as picking

a smaller value for G or collapsing neighboring groups. We also implemented an estimator that collapses the empty cell with a neighboring cell as well as a version with a lower bound on the value of the denominator in the weighting adjustment (*i.e.*, $\hat{t}^{*}_{\mathrm{WC}}$), and the results are virtually indistinguishable from those reported below, so they will not be further discussed here.

4.2 Results

Table 2 and 3 show the simulated bias of the weighting cell estimator for the variables Y_1 and Y_2 as a fraction of the standard deviation. As a comparison, the last column of Tables 2 and 3 displays the bias of the naive estimator, $\bar{y}_r$. The bias as a fraction of the standard deviation, referred to here as the *relative bias*,

$$\mathrm{RB}(\hat{t}_{\mathrm{WC}},\ \hat{t}_y) = \frac{\mathrm{E}(\hat{t}_{\mathrm{WC}} - \bar{Y})}{(\mathrm{Var}(\hat{t}_{\mathrm{WC}}))^{1/2}}$$

was also used in Cochran (1977, page 14), where it is shown that as the relative bias increases, inferential results rapidly become unreliable. In a simple simulation example, Cochran (1977) shows that a relative bias of ± 0.50 or more leads to highly inaccurate 95% confidence intervals.

For Y_1 (Table 2), the relative bias of the weighting cell estimator is small and is similar to the relative bias of the naive estimator, for all sample sizes, response mechanisms and cells sizes considered. For the variable Y_2 (Table 3), similar results hold when the response mechanism is uniform (C1, C2). However, when the response probabilities are a linear function of X (L1, L2), the naive estimator becomes severely biased. This relative bias decreases as the number of cells increases, and three to five cells appear sufficient to remove most of the bias. This finding agrees with that of Cochran (1968) in the context of bias reduction for observational studies.

Table 2
Relative Bias of the Weighting Cell and Naive Estimators for the Mean Y_1

Sample size	Response mechanism	Number of Cells 2	3	5	8	Naive estimator
200	C1	−0.00	−0.01	0.01	0.01	−0.00
	C2	0.01	−0.00	−0.01	0.00	0.00
	L1	−0.02	0.03	−0.04	−0.01	−0.00
	L2	−0.00	−0.02	0.00	−0.02	−0.00
500	C1	−0.00	−0.01	0.04	−0.01	0.00
	C2	0.01	0.02	−0.01	−0.01	0.00
	L1	0.05	0.02	−0.01	−0.02	0.01
	L2	0.01	0.01	−0.00	−0.01	0.01

Table 3
Relative Bias of the Weighting Cell and Naive Estimators for the Mean of Y_2

Sample size	Response mechanism	Number of Cells 2	3	5	8	Naive estimator
200	C1	0.01	−0.01	−0.02	0.02	−0.01
	C2	−0.03	−0.00	0.02	0.01	−0.00
	L1	1.16	0.59	0.22	0.07	3.57
	L2	0.36	0.18	0.06	0.03	1.36
500	C1	0.01	0.01	−0.02	−0.00	0.00
	C2	0.02	−0.00	−0.00	−0.01	−0.01
	L1	1.98	0.96	0.32	0.15	5.84
	L2	0.61	0.29	0.09	0.02	2.26

Hence, when the variable of interest is totally unrelated to the response mechanism, as in the cases of Y_1 under all mechanisms considered and of Y_2 under the uniform response mechanism, the bias does not depend on the number of cells. When the variable of interest and the response mechanism are related, multiple cells are required to remove the bias.

The relative mean squared error (RMSE) for the two variables of interest, defined as the MSE of the weighting cell estimator divided by the MSE of the estimator with no non-response,

$$\mathrm{RMSE}(\hat{t}_{\mathrm{WC}},\ \hat{t}_y) = \frac{\mathrm{E}(\hat{t}_{\mathrm{WC}} - t_y)^2}{\mathrm{E}(\hat{t}_y - t_y)^2},$$

are in Tables 4 and 5. In these tables, the last column again corresponds to the relative MSE of the naive estimator. Note that with the exception of the two L1 cases for variable Y_2, the Tables 4 and 5 are really variance tables, since the bias is so small.

For Y_1 (Table 4), the variable uncorrelated with X, the number of cells has relatively little effect on the relative mean square error, with results around 2.3 for a 50% response rate, and around 1.3 for the 80% rate. However, a relatively modest increase in MSE is observed, especially for the high nonresponse cases (C1, L1). For Y_2 (Table 5), the variable correlated with X, increasing the number of cells improves the results for all response mechanisms, but the effect is much more pronounced when the response mechanism is also correlated with the variable of interest. As for the relative bias, three to five cells achieve most of the efficiency gain, while the naive estimator is extremely inefficient.

Table 4
Relative Mean Squared Error of the Weighting Cell Estimator Compared to the Estimator Without Nonresponse for Y_1

Sample size	Response mechanism	Number of Cells 2	3	5	8	Naive estimator
200	C1	2.02	2.13	2.11	2.21	2.08
	C2	1.25	1.31	1.29	1.28	1.28
	L1	2.34	2.32	2.61	2.70	2.08
	L2	1.30	1.29	1.29	1.31	1.28
500	C1	2.25	2.21	2.19	2.31	2.23
	C2	1.30	1.32	1.34	1.29	1.30
	L1	2.55	2.57	2.62	2.70	2.22
	L2	1.32	1.35	1.33	1.34	1.31

Table 5
Relative Mean Squared Error of the Weighting Cell Estimator Relative to the Estimator Without Nonresponse for Y_2

Sample size	Response mechanism	Number of Cells 2	3	5	8	Naive estimator
200	C1	1.33	1.17	1.10	1.07	2.07
	C2	1.09	1.05	1.02	1.02	1.26
	L1	3.14	1.57	1.16	1.12	26.32
	L2	1.23	1.07	1.03	1.01	3.57
500	C1	1.35	1.19	1.10	1.09	2.22
	C2	1.09	1.05	1.03	1.03	1.30
	L1	6.60	2.30	1.23	1.13	69.75
	L2	1.50	1.14	1.04	1.02	7.83

The difference between the results for both variables is surprising at first, but it can be explained using the results from section 3. Clearly, the results for Y_2 follow the asymptotic theory, in that the MSE improves as the number of cells improves (as long as sufficient observations are available in each cell). In the case of Y_1, note first that the bias is negligible relative to the standard deviation for all values of G (see Table 2), so that the change in MSE is due almost exclusively to differences in variance. It turns out that when a variable is iid in the population and sampling is equal-probability, the asymptotic variance in Theorem 3.1 is relatively insensitive to the number of cells. In that case, the increase in MSE is influenced by the variability implied in the linear approximation in Theorem 3.1, which increases with the number of cells.

The theory described in this article applies to response functions that can have arbitrary smooth shape. In order to evaluate results for more complicated functions, we also created a variable $Y_3 = 25 + 95X - 95X^2 + \varepsilon$, where $\varepsilon \sim N(0,3)$, so that the Y_3 has mean 40.9 and variance 51.8, and two additional quadratic response mechanisms

$$-\varphi_{Q1}(X) = 0.17 + 1.96X - 1.96X^2$$

$$-\varphi_{Q2}(X) = 0.50 + 1.80X - 1.80X^2.$$

The results (not shown) broadly reflect the findings for the previous variables. When the response mechanism and the variables are correlated (the linear variable is correlated with the linear response mechanism, and the quadratic variable is correlated with the linear and quadratic response mechanisms), significant bias occurs but can be removed by increasing the number of cells. In the case of the quadratic response mechanism and the quadratic variable, eight or more cells appear to be required to remove the bias. Similarly, the relative efficiency improves for all response mechanisms for both the linear (Y_2) and quadratic variable, with the most dramatic results found for the linear variable/linear response and quadratic variable/quadratic response cases.

In the previous sections of this article, we approximated the weighting cell estimator by a "linearized" estimator $\tilde{t}_{\mathrm{WC}}$, and then derived the asymptotic properties of that estimator. It is therefore of interest to compare the statistical properties of both estimators in simulated settings. For all the scenarios in Table 1, we calculated the relative efficiencies of the weighting cell estimator compared to the linearized estimator. These relative efficiencies were all close to 1.00, with the largest deviation being a value of 1.08. Hence, the statistical properties of weighting cell estimator appear to be well approximated by those of the linearized estimator.

5. Conclusions

We have shown that the weighting cell estimator, corresponding also to the FEFI estimator proposed by Kim and Fuller (1999), is consistent with respect to the sampling design and a nonparametric response model. That model does not require the correct specification of homogeneous response probability cells, as long as a variable related to the response probability can be identified.

The statistical properties of the estimator depend on the number of cells used in the estimation, but the relationship is rather complex. Asymptotically, there appears to be a trade-off between the goodness of the approximation of the weighting cell estimator by a linearized estimator, which requires a small number of cells, and the mean squared error of that linearized estimator, which is reduced when a large number of cells are used. While useful in understanding the asymptotic behavior of the estimator, these

findings only provide limited guidance for choosing the number of cells for a particular survey. However, these findings show that reliable inference for weighting cell estimators will require cells with reasonable sample sizes, because variance estimates typically rely on the variance of the linearized estimator as an approximation of the variance of the weighting cell estimator.

The simulation experiments show that when the variable of interest and the response mechanism are uncorrelated, the number of cells has virtually no effect on the design bias of the estimator. When the variable of interest and the response mechanism are uncorrelated, even the estimator with a single weighting cell (corresponding to a simple ratio adjustment) is essentially unbiased, while models with multiple cells perform equally well. When the response mechanism and the variable of interest are related, however, the bias properties of the weighting cell estimator depend critically on the number of cells. In particular, estimators with a single cell are severely biased, but even a relatively small number of cells is sufficient to reduce both the bias and variance of the estimator. This result holds for both linear and nonlinear relationships between the response mechanism and the variable of interest.

The design efficiency of estimators depends on the relationship between the variable of interest and the variable(s) used to form weighting cells. When those two variables are uncorrelated, the number of cells has no effect on the efficiency of the estimator. Conversely, when those two variables are correlated, increasing the number of cells improves the design efficiency of the estimator. Even a small number of cells dramatically improves the performance of the estimator.

Overall, it appears that in the presence of nonresponse, forming at least a small number of weighting cells based on a variable related to the non-response provides a good "insurance policy" against design bias and design inefficiency. This article has shown that this adjustment does not require the assumption that the cells be based on *a priori* knowledge of constant nonresponse groups. The resulting weighting cell estimator will never perform worse than the naive estimator with a single ratio adjustment for the whole sample, and it might perform significantly better.

6. Acknowledgements

The authors thank Wayne Fuller for many helpful comments made during the development of this manuscript. We also are grateful for the comments of the associate editor and the two referees. This research was supported by a subcontract between Westat and Iowa State University under Contract No. ED-99-CO-0109 between Westat and the U.S. Department of Education. The first author gratefully acknowledges the support of CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil, during his Ph.D. studies at Iowa State University.

Appendix

Derivations of Theoretical Results

Lemma 1. *Assume that the conditions* $(A1)-(A3)$ *and* $(R1)-(R2)$ *hold. For* $i_1, i_2, ..., i_k \in U_v$, *define*

$$\Gamma_{i_1,\ldots,i_k} = \mathrm{E}\left(\prod_{l=1}^{k} (I_{i_l} R_{i_l} - \pi_{i_l} \varphi_{i_l}) \right),$$

where $\varphi_i = \varphi(X_i)$. *Consider the* $\Delta_{i_1,\ldots,i_k}$ *of* (3). *Let* A^r *denotes the* $r-$fold *Cartesian product of the set* A, *where* r *is a fixed positive integer,* $A_{1,r,v} = \{(i_1, i_2, ..., i_r) \in U_v^r : i_1 = i_2 = ... = i_r\}$ *and* $A_{k,r,v} = \{(i_1, i_2, ..., i_r) \in U_v^r :$ *exactly* k *components are distinct*$\}$, $k = 2, 3, ..., r$. *Then, for* $r = 8$,

$$N_v^8 n_v^{-8} \max_{i_1,\ldots,i_g \in A_{k,8,v}} (|\Gamma_{i_1,\ldots,i_8}|, |\Delta_{i_1,\ldots,i_8}|) = \begin{cases} O(N_v^3 n_v^{-4}), \text{if } k = 5 \\ O(N_v^3 n_v^{-5}), \text{if } k = 6 \\ O(N_v n_v^{-4}), \text{if } k = 7 \\ O(n_v^{-4}), \quad \text{if } k = 8. \end{cases}$$

Proof of Lemma 1. See Da Silva (2003).

Lemma 2. *Suppose the conditions of Theorem* 3.1 *hold. Consider the vectors* $\hat{t}_{gv} = (\hat{t}_{1,g}, \hat{t}_{2,g}, \hat{t}_{3,g})' = \sum_{U_g} \pi_i^{-1} (1, Y_i R_i, R_i)' I_i$ *and* $\hat{t}_{gv}^* \equiv (\hat{t}_{1,g}, \hat{t}_{2,g}, \hat{t}_{3,g}^*)'$, *with* $\hat{t}_{3,g}^* = \max\{\hat{t}_{3,g}, N_g G_v / n_v\}$. *Let* $\boldsymbol{t}_{gv} = E(\hat{\boldsymbol{t}}_{gv})$. *Then for all* $g = 1, 2, ..., G_v$,

$$\frac{1}{N_g^8} (\mathrm{E} \| \hat{\boldsymbol{t}}_{gv}^* - \boldsymbol{t}_{gv} \|^8, \mathrm{E} \| \hat{\boldsymbol{t}}_{gv} - \boldsymbol{t}_{gv} \|^8) = O((G_v / n_v)^4).$$

Proof of Lemma 2: See Da Silva (2003).

Proof of Theorem 3.1: Consider the proof of (5). Let $\boldsymbol{a} = (a_1, a_2, a_3)' \in \mathrm{R}^3$ and $h: \mathrm{R}^3 \to \mathrm{R}$, where $h(\boldsymbol{a}) = a_1 a_2 / a_3$, $a_3 \neq 0$. Define

$$\eta_{gv}(\boldsymbol{a}) = h(N_g^{-1} \boldsymbol{t}_{gv}) + \sum_{k=1}^{3} h^{(k)}(N_g^{-1} \boldsymbol{t}_{gv})(a_k - N_g^{-1} \boldsymbol{t}_{gv}),$$

where $h^{(k)}(\boldsymbol{a}) = \partial h(\boldsymbol{a}) / \partial a_k$, and let $e_{gv} = h(\boldsymbol{a}) - \eta_{gv}(\boldsymbol{a})$. Note that $\hat{t}_{\mathrm{WC}}^* = \sum_{g=1}^{G_v} N_g h(N_g^{-1} \hat{\boldsymbol{t}}_{gv}^*)$, and hence, defining the "linearized" estimator $\tilde{t}_{\mathrm{WC}} = \sum_{g=1}^{G_v} N_g \eta_{gv}(N_g^{-1} \hat{\boldsymbol{t}}_{gv})$, we can write

$$\frac{1}{N_v}(\hat{t}_{\mathrm{WC}}^* - \tilde{t}_{\mathrm{WC}}) = \bar{e}_v + \bar{\eta}_v,$$

where

$$\bar{e}_v = \frac{1}{N_v}\sum_{g=1}^{G_v} N_g\, e_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)$$

and

$$\bar{\eta}_v = \frac{1}{N_v}\sum_{g=1}^{G_v} N_g\,(\eta_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*) - \eta_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv})).$$

Consider first the term $\bar{\eta}_v$. Observe that

$$|\eta_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*) - \eta_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv})| = |h^{(3)}(N_g^{-1}\boldsymbol{t}_{gv})|\frac{1}{N_g}|\hat{t}_{3,g}^* - \hat{t}_{3,g}|.$$

By (A4) and (A5), it is straightforward to check that $h(\cdot)$ and $h^{(k)}(\cdot)$, $k = 1, 2, 3$, are O(1) when evaluated at $N_g^{-1}\boldsymbol{t}_{gv}$, for all $g = 1, 2, ..., G_v$. Since by construction, we have $1/N_g\,|\hat{t}_{3,g}^* - \hat{t}_{3,g}| \le G_v/n_v$, we conclude that $|\bar{\eta}_v| = O(G_v/n_v)$. Thus, to complete the proof of (5), it remains to show that $\bar{e}_v = O_p(G_v n_v^{-1})$. Let $f_{gv}(\boldsymbol{a}) \equiv (e_{gv}(\boldsymbol{a}))^2$. By the C_r the inequality (Sen and Singer 1993, page 21),

$$|f_{gv}(\boldsymbol{a})|^2 \le 5^3\left(\begin{array}{l}|h(\boldsymbol{a})|^4 + |h(N_g^{-1}\boldsymbol{t}_{gv})|^4 \\ \quad + \sum_{k=1}^{3}|h^{(k)}(N_g^{-1}\boldsymbol{t}_{gv})|^4\,|a_k - N_g^{-1}\boldsymbol{t}_{gv}|^4\end{array}\right).$$

Using (A1) and (A4), straightforward bounding arguments show that $|h(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)|^4 = O((n_v/G_v)^4)$ and that $N_g^{-4}|\hat{t}_{k,g} - t_{k,g}|^4 = O(1)$ for $k = 1, 2, 3$. Therefore,

$$|f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)|^2 = O\left(\frac{n_v^4}{G_v^4}\right).$$

Since by Lemma 2, $N_g^{-8}\,E\,\|\hat{t}_{gv}^* - \boldsymbol{t}_{gv}\|^8 = O((n_v/G_v)^{-4})$, and $v\,|f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)|^2$ is continuous at any realization of $N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*$, then the sequence $\{|f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)|^2\}$ satisfies the conditions of Theorem 5.4.4 (with $\eta = 1$, $p = 4$) of Fuller (1996, page 247). Therefore,

$$\mathrm{E}[|f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)|^2] = O(1),\ \forall g = 1, 2, ..., G_v.$$

Now, from the continuity of $f_{gv}(\cdot)$ and its derivatives up to order three, $\{f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)\}$ satisfies the conditions of Theorem 5.4.3 (with $\delta = 1$, $s = 4$ and $a_v = O(\sqrt{G_v/n_v})$ of Fuller (1996, pages 244 – 245). Hence,

$$\mathrm{E}f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*) = O(a_v^4) = O\left(\frac{G_v^2}{n_v^2}\right),\ \forall g = 1, 2, ..., G_v,$$

because $f_{gv}(\cdot)$ and all of its derivatives up to order three are zero at $N_g^{-1}\boldsymbol{t}_{gv}$. Therefore, we conclude that

$$\mathrm{E}|\bar{e}_v| \le \frac{1}{N_v}\sum_{U_g} N_g\,\mathrm{E}|e_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*)| \le \frac{1}{N_v}\sum_{U_g} N_g\,(\mathrm{E}f_{gv}(N_g^{-1}\hat{\boldsymbol{t}}_{gv}^*))^{1/2} = O\left(\frac{G_v}{n_v}\right),$$

which leads to $\bar{e}_v = O_p(G_v n_v^{-1})$ by an application of Markov's inequality.

Expressions (6) and (7) are obtained by direct computation of the moments of the linear estimator $\tilde{t}_{\mathrm{WC}}$ under the sampling design and the response mechanism.

Proof of Corollary 3.1: Let

$$Z_v = \frac{1}{V_v^{1/2}}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v} - \bar{Y}_v - B_v\right)$$

and

$$W_v = \frac{1}{V_v^{1/2}}\left(\frac{\hat{t}_{\mathrm{WC}}^*}{N_v} - \frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right),$$

where $V_v = \mathrm{Var}(\tilde{t}_{\mathrm{WC}}/N_v)$. Hence,

$$\left[\mathrm{Var}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right)\right]^{1/2}\left(\frac{\hat{t}_{\mathrm{WC}}^*}{N_v} - \bar{Y}_v - B\right) = Z_v + W_v.$$

Since $V/n_v V_v \to 1$, as $v \to \infty$, then,

$$Z_v = \frac{1}{V^{1/2}}\left(\frac{V}{n_v V_v}\right)^{1/2} n_v^{1/2}\left(\frac{\tilde{t}_{\mathrm{WC}}}{N_v} - \bar{Y}_v - B_v\right) \underset{L}{\to} \frac{1}{V^{1/2}}Z,$$

where $Z \sim N(0, V)$. Also, (A7) with $\gamma < 1/2$ implies that $n_v^{1/2}O_p(G_v n_v^{-1}) = o_p(1)$. Hence, by Theorem 3.1,

$$W_v = \frac{1}{V^{1/2}}\left(\frac{V}{n_v V_v}\right)^{1/2} n_v^{1/2}\left(\frac{\hat{t}_{\mathrm{WC}}^*}{N_v} - \frac{\tilde{t}_{\mathrm{WC}}}{N_v}\right) = o_p(1).$$

The result of the corollary follows, therefore, from Fuller (1996, Theorem 5.2.1).

Proof of Theorem 3.1: Fix a $g \in \{1, 2, ..., G_v\}$. The conditions of the theorem imply, by the Intermediate Value Theorem, that there exists X_{0g} inside the interval defined by the lowest and the highest values of $X_i \in U_g$ such that $\bar{\varphi}_g = N_g^{-1}\sum_{U_g}\varphi_i = \varphi(X_{0g})$. Also, by the mean Value Theorem, $\forall i \in U_g$,

$$\varphi_i = \varphi(X_i) = \varphi(X_{0g}) + \varphi'(c^*)(X_i - X_{0g}),$$

where c^* is between X_i and X_{0g}. So,

$$|\varphi_i - \bar{\varphi}_g| = |\varphi'(c^*)\|X_i - X_{0g}| \le C\frac{X_{(N)} - X_{(1)}}{G_v}, \quad (8)$$

for some constant $C \in (0, \infty)$ and, by (A5) and (A6),

$$\left| \text{Bias}\left(\frac{\tilde{t}_{\text{WC}}}{N_v} \right) \right| \leq C\lambda_6^{-1}\lambda_5 \frac{X_{(N)} - X_{(1)}}{G_v}.$$

Observe now that since

$$\left| \tilde{Y}_{ig} \right| \leq \frac{1}{\pi_i}\frac{\varphi_i}{\overline{\varphi}_g}|Y_i| + \frac{1}{\pi_i}\frac{\left|\varphi_i - \overline{\varphi}_g\right|}{\overline{\varphi}_g}\left|\tilde{Y}_g\right|,$$

then, by (A1, A6) and (8),

$$\tilde{Y}_{ig} = O\left(\frac{N_v}{n_v}\right) + O\left(\frac{N_v}{n_v G_v}\right), \quad \forall U_g, \forall g = 1, 2, ..., G_v,$$

which implies that

$$\tilde{Y}_{ig}\tilde{Y}_{jg'} = O\left(\frac{N_v^2}{n_v^2}\right) + O\left(\frac{N_v^2}{n_v^2 G_v}\right), \forall U_g, \forall g = 1, 2, ..., G_v.$$

Using the facts that, by (A7), $N_g / N_v = O(1/G_v)$, by (A2) and (A3), $\sum_{U_g}\sum_{U_g}\Delta_{ij} = O(n_v / G_v)$ and, for $g \neq g'$, $\sum_{U_g}\sum_{U_{g'}}\Delta_{ij} = O(n_v / G_v^2)$, then, the first term of $\text{Var}(\tilde{t}_{\text{WC}} / N_v)$ is bounded by

$$O\left(\frac{1}{n_v}\right) + O\left(\frac{1}{nG_v}\right).$$

Since the second terms of $\text{Var}(\tilde{t}_{\text{WC}} / N_v)$ is bounded by $O(1/n_v)$, the conclusion follows.

References

Cassel, C.-M., Särndal, C.-E. and Wretman J.H. (1983). Some uses of statistical models in connection with the nonresponse problem. In *Incomplete data in sample surveys*: *Theory and bibliographies*, (Eds. W.G. Madow, I. Olkin and D. B. Rubin). Academic Press, New York: London. 3, 143-160.

Cochran, W.G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 24, 295-313.

Cochran, W.G. (1977). *Sampling Techniques* (3rd Ed.). New York: John Wiley & Sons, Inc.

Da Silva, D.N. (2003). Adjustments for Survey Unit Nonresponse Under Nonparametric Response Mechanisms. Ph. D. Thesis, Iowa State University, Ames, IA.

Da Silva, D.M., and Opsomer, J.D. (2003). A kernel smoothing method to adjust for unit nonresponse in sample surveys. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association [CD-ROM]. Alexandria, V.A. Article #00605.

Fuller, W.A. (1996). *Introduction to Statistical Time Series* (Second Edition). Wiley.

Fuller, W.A., and Kim, J.-K. (2003). Hot deck imputation for the response model. Submitted for publication.

Hansen, M.H., Madow, W.G. and Tepping, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Kalton, G. (1983). *Compensating for Missing Survey Data*. Institute of Social Research.

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.

Kalton,G., and Maligalig, D.S. (1991). A comparison of methods of weighting adjustment for nonresponse. In *Proceedings of the Bureau of the Census Annual Research Conference*. U.S. Bureau of the Census (Suitland, MD). 409-428.

Kim, J.-K. and Fuller, W.A. (1999). Jackknife variance estimation after hot deck imputation. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association. Alexandria, VA, 825-830.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157.

Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis With Missing Data*. Wiley, 20.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustments for unit non-response. In *Incomplete data in sample surveys: Theory and bibliographies*, (Eds. W.G. Madow, I. Olkin, and D.B. Rubin). Academic Press New York: London. 2, 143-184.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Sen, P.K., and Singer, J.D.M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman and Hall Ltd.

U.S. Bureau of the Census (1963). The Current Population Survey: A report on Methodology. Technical Paper No. 7, Washington, DC.